Machine Learning in Macroeconomic Forecasting: Enhancing Predictions of GDP, Inflation, and Recessions

George Mathew Keluthara

Abstract— Macroeconomic forecasting has normally relied on linear econometric models such as ARIMA and SARIMA. However, these previous models usually struggle to capture nonlinear relationships and high-dimensional data that is present in modern economics. This research paper will explore the application of machine learning techniques such as Random Forests, Gradient/Boosting, Lasso regression as well as Recurrent Neural Networks (RNN/LSTM), to improve the prediction of macroeconomic information like GDP growth, inflation, and recession risks. Drawing on recent empirical research, this research paper demonstrates that machine learning models can outperform traditional methods, reducing forecasting error and improving the responsiveness to macroeconomic shocks. The findings support the integration of machine learning into macroeconomic modeling while also acknowledging its limitations, such as interpretability and overfitting.

Index Terms— macroeconomic forecasting, machine learning, GDP, inflation, recession prediction.

I. INTRODUCTION

Accurate macroeconomic forecasting is essential for policymakers, investors and also institutions. Traditional forecasting models, such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant the Seasonal Autoregressive Integrated Moving Average (SARIMA), have been used to predict key indicators like GDP growth, inflation and also unemployment. However, these models rely mostly on linear assumptions as well as limited datasets, which limit their ability to capture the complexity of modern economics.

Recent advancements in machine learning offer new tools that are capable of handling nonlinear relationships, large datasets as well as real-time information. This paper will investigate how machine learning models can enhance forecasting key macroeconomic indicators, focusing on three key areas: GDP growth, inflation, and recession prediction.

II. LITERATURE REVIEW

A. GDP Growth Prediction

Recent studies demonstrate that machine learning models can outperform traditional econometric approaches in predicting GDP growth. Techniques such as Random Forest and Gradient Boosting leverage decision tree ensembles to capture nonlinear relationships among economic variables in ways that ARIMA-family models cannot. Research on Brazil and U.S. GDP forecasting shows that tree-based models significantly reduce prediction error compared to ARIMA-based approaches, incorporating large datasets – including financial indicators, industrial production, and trade variables – for more accurate and timely predictions.

A key study by Kant et al. (2024) examined GDP nowcasting in the Netherlands using a dataset of 83 macroeconomic and financial predictors over the period 1992Q1–2018Q4. The authors found that Random Forest provided the most accurate nowcasts on average, outperforming the Dynamic Factor Model (DFM) – the most widely used conventional nowcasting benchmark – particularly over the near-term horizon. The study also employed Shapley values to explain which predictors drove the model's forecasts, offering a bridge between predictive power and transparency.

Kim et al. (2025) further developed this work by comparing RNN, LSTM, GRU, Temporal Convolutional Networks (TCN), and Transformer models for per capita GDP prediction. Incorporating the Consumer Price Index (CPI) and unemployment rate as additional predictors, the Transformer model consistently achieved the highest R-squared and lowest error values across all metrics – suggesting attention-based architectures may represent the next frontier in GDP forecasting.

Research on China's economy similarly found that ML models generally produce lower average forecast errors than traditional econometric models or expert predictions, with performance advantages most pronounced during periods of economic stability. Interpretable ML methods identified industrial production, the Purchasing Managers' Index (PMI), and retail sales as among the most informative predictors.

LASSO Regression has emerged as a valuable complementary technique, applying L1 regularization to perform automatic variable selection in high-dimensional settings. Goulet Coulombe et al. (2019) confirm that LASSO and related regularization methods are valuable for managing large predictor sets in macroeconomic modeling, though they remain bounded by their linearity assumption and do not universally outperform factor models.

B. Inflation Prediction

Inflation forecasting presents a distinct set of challenges. Inflation dynamics are shaped by monetary conditions, supply-side shocks, demand pressures, exchange rate movements, and expectations – many of which interact nonlinearly. Traditional approaches such as the Augmented Phillips Curve, ARIMA, and VAR models frequently fail to capture these interactions, leading to systematic forecast errors, particularly around structural breaks and periods of economic uncertainty.

Machine learning has gained considerable momentum in this area. Medeiros et al. (2021) applied a range of ML methods to predict U.S. inflation between one and twelve months ahead, finding that Random Forest consistently delivered the most precise forecasts. This finding has been replicated across multiple subsequent studies, making Random Forest one of the most empirically validated ML approaches for inflation prediction.

Research on emerging economies further demonstrates the value of ML methods. A 2024 ScienceDirect study examined inflation forecasting across a panel of emerging market countries, incorporating foreign exchange (FX) reserves as a novel predictor.

The results showed that Random Forest and Gradient Boosting consistently outperformed traditional econometric models across all sets of determinants, with the inclusion of FX reserves improving accuracy for both traditional and ML models alike.

A significant methodological advance visible in the recent literature is the shift from point forecasting to distributional forecasting – predicting the full probability distribution of inflation outcomes rather than a single expected value. This approach provides central banks with richer information about upside and downside risks, aligning well with the risk management mandate of monetary authorities.

It is important to note, however, that ML methods are not uniformly superior. Goulet Coulombe et al. (2019) find that nonlinearity actually reduces forecasting accuracy for U.S. inflation using standard datasets, while benefiting unemployment and industrial production forecasting. This underscores a critical principle: ML advantages are variable-specific and context-dependent, and practitioners must validate carefully before adopting ML models for any specific application.

C. Recession and Economic Crisis Prediction

Predicting the onset of economic recessions and financial crises represents arguably the highest-stakes application of macroeconomic forecasting. Recessions are rare events characterized by structural breaks and nonlinear dynamics that are poorly captured by linear models trained predominantly on expansionary periods, yet their social and economic costs make early warning capability enormously valuable.

Wang et al. (2021) developed a Bi-LSTM with Autoencoder architecture to forecast U.S. recessions, demonstrating early warning capabilities – predicting recessions six months ahead of their official NBER dates – and performing well across both the Great Financial Crisis and the COVID-19 recession. The latter result is especially significant, as the COVID-19 recession was triggered by an exogenous health shock with no close historical parallel.

A 2024 arXiv working paper applied LSTM and GRU models to US recessions from 1967 to 2021, using SHAP decomposition to identify key drivers. The S&P 500 index emerged as the most important predictor for short-term (up to three-month) forecasts, while the Treasury term spread dominated longer-term (up to twelve-month) predictions. Both LSTM and GRU models outperformed standard probit and logit benchmarks while also capturing business cycle asymmetries that linear models inherently cannot represent.

Barthélémy (2024) developed an early warning system for currency crises across 68 countries over the period 1995–2020. LSTM and GRU models outperformed logistic regression and Random Forest benchmarks, producing fewer false alarms and more continuous alert signals. At a longer horizon, Tölö (2020) found that when calibrated to detect more than 80% of crises, an LSTM model produced less than half the number of false alarms relative to a logit benchmark – a major practical advantage given the social cost of unnecessary interventions triggered by false alarms.

D. . Methodology

E. Research Design

This paper is a systematic literature review and comparative analysis of empirical methodologies in ML-based macroeconomic forecasting. It synthesizes findings from peer-reviewed articles, working papers, and institutional reports published between 2018 and 2026, prioritizing studies that directly compared ML models to traditional econometric benchmarks on the same datasets, focused on GDP, inflation, or recession prediction, and employed rigorous out-of-sample evaluation frameworks.

F. Machine Learning Models Reviewed

Ensemble Tree-Based Methods. Random Forest constructs a large number of decision trees on bootstrapped subsamples of data, averaging their predictions to reduce variance. It naturally handles high-dimensional datasets, does not require stationarity, and captures nonlinear and interaction effects. Gradient Boosting builds trees sequentially, each correcting the residual errors of its predecessor, and its implementations – XGBoost and LightGBM – are among the most competitive forecasting tools in both academic and applied settings.

Regularized Regression. LASSO applies L1 regularization to shrink irrelevant coefficients to zero, performing automatic variable selection. Ridge regression and Elastic-net offer related approaches with different selection properties. These methods are valuable for managing large predictor sets but remain bounded by their linearity assumption.

Recurrent Neural Networks (LSTM and GRU). LSTM networks were designed to model long-range temporal dependencies in sequential data, making them naturally suited to the lagged, path-dependent dynamics of economic cycles. GRU is a computationally lighter variant offering comparable performance. Bidirectional LSTM (Bi-LSTM) processes sequences in both forward and reverse temporal directions, further improving recession prediction accuracy.

G. Evaluation Framework

The primary evaluation metrics across the reviewed literature include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Diebold-Mariano test for comparing model pairs. For recession and crisis prediction, studies use accuracy, sensitivity, specificity, the Area Under the ROC Curve (AUC), and false alarm rates. The most rigorous studies employ pseudo-out-of-sample (POOS) or rolling window evaluation frameworks, which mimic real-time forecasting conditions and guard against in-sample overfitting.

H. . Results and Discussion

I. GDP Forecasting

ML models demonstrate consistent improvements in GDP forecasting accuracy, particularly in nowcasting – where the objective is to estimate current-quarter GDP growth before official data is released. Random Forest models exploiting large panels of monthly indicators significantly outperform the Dynamic Factor Model in this context (Kant et al., 2024). During periods of elevated economic volatility, ML models offer their greatest relative advantage, though research on China finds the margin narrows at economic inflection points – suggesting that expert judgment retains complementary value at turning points where structural context matters most. This implies ML tools should be integrated with domain expertise rather than used as wholesale replacements for economic judgment.

The inclusion of unconventional high-frequency data – from financial markets, logistics, and geospatial sources – further enhances ML model performance beyond what is achievable with standard macroeconomic datasets alone, illustrating the scalability of ML approaches to real-world institutional forecasting.

J. Inflation Forecasting

The inflation forecasting literature reveals a more nuanced picture. While Random Forest consistently outperforms linear benchmarks in point forecasts across the U.S., euro area, and emerging market economies, relative performance is sensitive to the economic environment and forecast horizon. The shift from point forecasting to distributional forecasting represents a significant advance – providing central banks with richer information about upside and downside risks that directly supports monetary policy risk management.

The finding that nonlinearity can actively harm U.S. inflation forecasting while helping unemployment and industrial production forecasting underlines a core lesson: ML methods are not universally superior, and careful empirical validation on a case-by-case basis is essential.

K. Recession and Crisis Prediction

Recession and crisis prediction is where deep learning demonstrates its most compelling advantages. The classification nature of the task plays to the strengths of neural network architectures, which can learn highly nonlinear decision boundaries and exploit temporal pattern recognition in ways probit and logit models cannot. Wang et al.'s (2021) six-month early warning result, and Tölö's (2020) finding that LSTM produces less than half the false alarms of logit models, have direct and significant implications for the early warning systems of central banks and international financial institutions. The SHAP-based interpretability results across multiple studies – identifying the term spread, equity indices, and real exchange rate deviations as the dominant predictors – provide reassurance that these models capture genuine economic relationships rather than spurious historical patterns.

L. . Limitations and Challenges

M. Interpretability

Traditional econometric models are structurally transparent: the relationship between each predictor and the outcome is explicit and theoretically grounded. Machine learning models – particularly LSTM – are functionally opaque. Post-hoc interpretability methods such as SHAP and LIME partially address this, but provide approximations of model behavior rather than true structural insight, and do not resolve whether the relationships learned are economically meaningful or historically contingent.

N. Data Requirements and Overfitting

Deep learning models require large training datasets to estimate their high-dimensional parameter spaces reliably. In macroeconomics, training data is typically limited – quarterly GDP data may cover only a few decades, and financial crises are rare enough that training samples are inherently small. Overfitting is a persistent risk that must be managed through cross-validation, regularization, and conservative model specification. Researchers should treat in-sample fit with skepticism and demand strong out-of-sample evidence before deploying ML models operationally.

O. Structural Breaks and Governance

Economic systems are subject to periodic structural breaks – fundamental changes in variable relationships driven by policy shifts, technological transformation, or extraordinary events like pandemics. ML models trained on historical data are vulnerable to distribution shift when these relationships change in ways not represented in training data. Traditional theory-grounded models may be more robust in such circumstances.

Additionally, as Klein et al. (2026) discuss, deploying ML forecasts in policy settings raises issues of accountability and explainability to democratic institutions. When central banks adjust interest rates partly based on opaque model outputs, robust institutional frameworks and professional standards become essential safeguards.

P. . Conclusion

The weight of evidence across the literature is clear: machine learning models – particularly Random Forests, Gradient Boosting, LASSO, and LSTM/GRU architectures – consistently outperform traditional econometric benchmarks, delivering lower forecast errors and earlier, more reliable early warnings of economic downturns. The demonstrated ability of LSTM to predict recessions six months in advance, and to generate substantially fewer false alarms in financial crisis prediction, suggests genuine potential for enhancing the early warning systems of central banks and international institutions.

However, the adoption of ML methods is not without significant challenges. The most productive path forward is to integrate ML models as complementary tools within hybrid frameworks that combine the pattern-recognition strengths of ML with the theoretical grounding and interpretability of conventional economic modeling – rather than positioning them as outright replacements.

Future research should prioritize the development of more robust interpretability methods, systematic evaluation of ML performance across different economic regimes, and institutional frameworks for the responsible deployment of ML tools in public policy. Realized thoughtfully, the integration of machine learning

into macroeconomic forecasting holds genuine promise for improving economic analysis and enabling better-informed decisions that affect the welfare of millions of people.

REFERENCES

[1] Barthélémy, P. (2024). Early warning system for currency crises using long short-term memory and gated recurrent unit neural networks. Journal of Forecasting. https://doi.org/10.1002/for.3069

[2] Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2019). How is machine learning useful for macroeconomic forecasting? University of Pennsylvania. https://economics.sas.upenn.edu/system/files/2019-03/GCLSS_MC_MacroFcst.pdf

[3] International Journal of Scientific Development and Research. (2024). An application of AI models in macroeconomic analysis. IJSDR, 9(12). https://ijsdr.org/papers/IJSDR2412051.pdf

[4] Kant, D., Pick, A., & Winter, J. D. (2024). Nowcasting GDP using machine learning methods. AStA Advances in Statistical Analysis. https://doi.org/10.1007/s10182-024-00515-0

[5] Kim, H., et al. (2025). Development of per capita GDP forecasting model using deep learning. Sustainability, 17(3), 843. https://doi.org/10.3390/su17030843

[6] Klein, L. C., et al. (2026). Macroeconomics' forecasting using machine learning approaches by policy makers. In Optimization, Learning Algorithms and Applications. Springer. https://doi.org/10.1007/978-3-032-00140-5_9

[7] Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment. Journal of Business & Economic Statistics, 39(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745

[8] Tölö, E. (2020). Predicting systemic financial crises with recurrent neural networks. Journal of Financial Stability, 49, 100746. https://doi.org/10.1016/j.jfs.2020.100746

[9] Wang, Z., Li, K., Xia, S. Q., & Liu, H. (2021). Economic recession prediction using deep neural network. ArXiv preprint arXiv:2107.10980. https://arxiv.org/abs/2107.10980

[10] Zhou, Y., et al. (2025). Enhancing economic cycle forecasting based on interpretable machine learning. Technological Forecasting and Social Change. https://doi.org/10.1016/j.techfore.2025.001258